

# On the identification of discrete graphical models with hidden nodes

Elena Stanghellini

D.E.F.S., Università di Perugia  
Via Pascoli 1, 06100 Perugia, Italy  
elena.stanghellini@stat.unipg.it

Barbara Vantaggi

Dept. Metodi e Modelli Matematici, Università “La Sapienza”  
Via Scarpa 16, 00161 Roma, Italy  
vantaggi@dmmm.uniroma1.it

## Abstract

Conditions are presented for local identifiability of discrete undirected graphical models with a binary hidden node. These models can be obtained by extending the latent class model to allow for conditional associations between the observed variables. We establish a necessary and sufficient condition for the model to be locally identified almost everywhere in the parameter space and we provide expressions of the subspace where identifiability breaks down. The condition is based on the topology of the undirected graph and relies on the faithfulness assumption.

**Keywords:** Conditional independence, contingency tables, finite mixtures, hidden variables, identifiability, latent class, log linear models

## 1 Introduction

In this paper we focus on undirected graphical models for discrete variables with one binary latent variable. These models generalize the latent class model, by allowing associations between the observed variables conditionally on the latent one. Allowing conditional associations between the observed variables in a latent class model is an alternative approach to add latent classes; see [14]. The practical importance of this issue is witnessed by several applied papers; see e.g. [7], [9] and [21].

In the recent literature the algebraic and geometric features of Bayesian networks with hidden nodes have been studied, see e.g. [8, 18, 19, 22]. As mentioned in these papers, when some of the variables are never observed, non-identifiability and local maxima in the likelihood function can occur and the dimension of the model is not easily computed. Furthermore, it has been shown in [10] that undirected discrete graphical models with no hidden variables are linear exponential family, while latent class models and more generally directed discrete graphical models with hidden nodes are members of the stratified exponential family, for which standard asymptotic results do not apply.

We concentrate on the identification of discrete undirected graphical models with one unobserved binary variable and establish a necessary and sufficient condition for the rank of the parametrization to be full. This ensures local identification of this class of models. The condition is based on the topology of the undirected graph associated to the model and relies on the faithfulness assumption. For non-full rank models, the obtained characterization allows us to find the expression of the (sub)space where the identifiability breaks down. Geometrically, this corresponds to the singularities in the parameter space [6, 10, 22].

The non-identifiability issue mentioned above has considerable repercussions on the asymptotic properties of standard model selection criteria (e.g. likelihood ratio statistic and other criteria, such as BIC), whose applicability and correctness may no longer hold. As stressed in [6, 10, 22] even when singularities are removed, the standard asymptotic tools may still be inappropriate for model selection of Bayesian networks with hidden nodes. In particular, in [22] an adjusted BIC score for naive Bayesian networks with one hidden variable is presented, with the correction depending on the types of singularities of the sufficient statistics of the postulated model. Large sample distributions of the likelihood ratio test with reference to singularities are studied in [6].

In Section 2 the model is presented, while in Section 3 the main derivations are detailed for the binary case. In Section 4, we extend the results to more complex models while in Section 5 we present our conclusions.

## 2 Identification of the discrete undirected graphical model

Let  $G^K = (K, E)$  be an undirected graph with node set  $K = \{0, 1, \dots, n\}$  and edge set  $E = \{(i, j)\}$  whenever vertices  $i$  and  $j$  are adjacent in  $G^K$ ,

$0 \leq i < j \leq n$ . To each node  $v$  is associated a discrete random variable  $A_v$ . A discrete undirected graphical model is a family of joint distributions of the variables  $A_v$ ,  $v \in K$ , satisfying the Markov property with respect to  $G^K$ , namely that  $A_v$  and  $A_u$  are conditionally independent given all the remaining variables whenever  $u$  and  $v$  are not adjacent in  $G^K$ ; see [13] (Ch. 3) for definitions and concepts.

Let  $A_0$  be a binary unobserved variable and  $O = \{1, \dots, n\}$  be the set of nodes associated to observed variables  $A_v$  with  $v \in O$ . In the following we let  $G^B$  to indicate the (sub)graph  $G^B = (B, E_B)$  of  $G^K$  induced by  $B \subseteq K$ . We denote with  $\bar{G}^B = (B, \bar{E}_B)$  the complementary graph of the (sub)graph  $G^B$ , where  $\bar{E}_B$  is the edge set formed by the pairs  $(i, j) \notin E_B$  with  $i, j \in B$  ( $i \neq j$ ). In Figure 1(b) the complementary  $\bar{G}^O$  corresponding to the graph  $G^K$  of Figure 1(a) is presented. We later prove that the corresponding graphical model is locally identified.

Let  $l_v$  indicate the number of levels of  $A_v$ ,  $v \in K$ , and let  $l = \prod_{v=1}^n l_v$ . Without loss of generality we assume that the variable  $A_v$  takes value in  $\{0, \dots, l_v - 1\}$ . We consider the multidimensional contingency table obtained by the cross classification of  $N$  objects according to  $A_v$ . Let  $X$  be the  $2l \times 1$  vector of entries of the contingency table, stacked in a way that  $A_0$  is running the slowest.

Data for contingency tables can be collected under various sampling scheme (see [13], Ch. 3). We assume for now that the elements of  $X$  are independent Poisson random variables with  $E(X) = \mu_X$ . Let  $\log \mu_X = Z\beta$ , where  $Z$  is a  $2l \times p$  design matrix defined in a way that the joint distribution of  $A_v$ ,  $v \in K$ , factorizes according to  $G^K$  and such that the model is graphical;  $\beta$  is a  $p$ -dimensional vector of unknown parameters. We adopt the corner point parametrization that takes as first level the cell with  $A_v = 0$ , for all  $v \in K$ , see e.g. [4]. We denote by  $Y$  the  $l \times 1$  vector of the counts in the marginal table, obtained by the cross classification of the  $N$  objects according to the observed variables only. The vector  $Y$  is stacked in a way that  $Y = LX$ , with  $L = (1, 1) \otimes I_l$ . By construction, the elements of  $Y$  are independent Poisson random variables with  $\mu_Y = Le^{Z\beta}$ .

Let  $P_Y(y, \cdot)$  be the joint probability distribution of  $Y$ . Let  $\Omega$  be the parameter space. A (parametric) model is *globally* identifiable if there exist no two distinct parameter values  $\beta, \beta^0 \in \Omega$  such that  $P_Y(y, \beta^0) = P_Y(y, \beta)$  (see [1]). A (parametric) model is *locally* identifiable in  $\beta^0$  if there exists an open neighborhood of  $\beta^0$  containing no other  $\beta \in \Omega$  such that  $P_Y(y, \beta^0) = P_Y(y, \beta)$ . If this condition is true for all  $\beta^0 \in \Omega$  the model is locally identified. Global identifiability of a model implies local identifiability, but not vice versa.

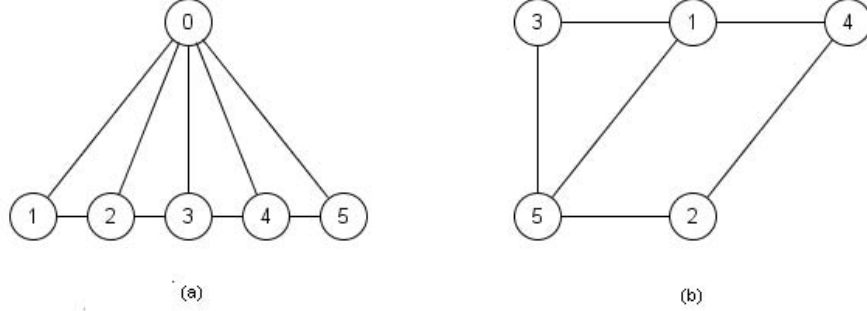


Figure 1: Example of (a) a  $G^K$  graph and (b) the corresponding graph  $\bar{G}^O$

We recall that, by the inverse function theorem, a model is locally identified if the rank of the transformation from the natural parameters  $\mu_Y$  to the new parameters  $\beta$  is full (see also [17, 2]). In this context, this is equivalent to the rank of following derivative matrix

$$D(\beta)^T = \frac{\partial \mu_Y^T}{\partial \beta} = \frac{\partial (Le^{Z\beta})^T}{\partial \beta} = (LRZ)^T \quad (1)$$

being full, where  $R = \text{diag}(\mu_X)$ . Note that the  $(i, j)$ -th element of  $D(\beta)$  is the partial derivative of the  $i$ -th component of  $\mu_Y$  with respect to  $\beta_j$  the  $j$ -th element of  $\beta$ . The multinomial case can be addressed in an analogous way to the Poisson, after noting that the rank of the matrix  $D(\beta)$  is equivalent to that of its submatrix  $D_0(\beta)$  obtained by deleting the last column.

Note that, by setting  $t_j = e^{\beta_j}$  for any parameter  $\beta_j$ , the parametrization map turns into a polynomial one. This implies, see e.g. [11, 16], that if there exists a point in the parameter space of  $t_j$ , and therefore on  $\Omega$ , at which the Jacobian has full rank, then the rank is full almost everywhere. Therefore, either there is no point in the parameter space at which the rank is full, or the rank is not full in a subspace of null measure. The object of this paper is (a) to establish a necessary and sufficient condition for the rank of  $D(\beta)$  to be full almost everywhere (b) to provide expressions of the subspace of null measure where identifiability breaks down.

### 3 Local identification with binary variables only

In this section we consider graphical models for binary variables only and assume that all  $n$  observed variables are connected to the unobserved one 0,

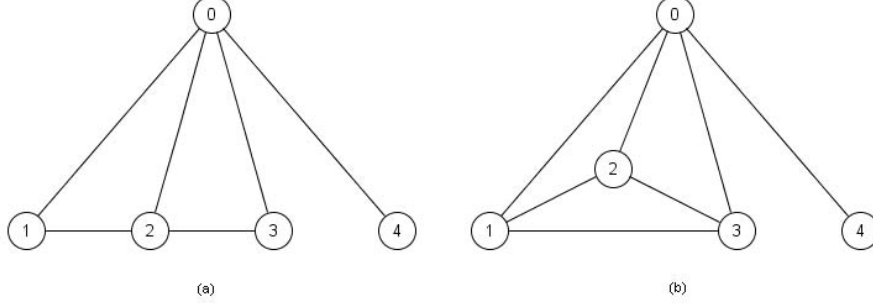


Figure 2: Two examples of  $G^K$  graphs

i.e.  $(u, 0) \in E$ , for any observed variable  $u \in O$ . Both assumptions will be relaxed in Section 5.

In this paper we assume that the graph is faithful. This implies that for each complete subgraph  $G^S = (S, E_S)$ ,  $S \subseteq O$ , (a) if  $|S| = 1$  there is a non-zero main effect of  $A_S$  and a non-zero second order interaction term between  $A_S$  and  $A_0$ ; (b) if  $|S| > 1$  there is a non-zero interaction term of order  $|S|$  among the variables in  $S$  and of order  $|S| + 1$  among the variables in  $\{0, S\}$ .

Let  $t$  be the maximum order of the non-zero interaction terms among the variables in  $O$ . For each order  $k$ ,  $k \in \{2, \dots, t\}$ , of interaction between the observed variables, let  $s_k$  be the number of interaction terms of order  $k$ . We use  $I_{k,r}$  to denote the set of vertices in  $O$  having a non-zero  $r$ -th interaction term of order  $k$ ,  $r \in \{1, \dots, s_k\}$ . Note that, by construction,  $|I_{k,r}| > 1$ . The following example clarifies the notation.

*Example 1.* The model with graph  $G^K$  as in Figure 2 (a) has maximum order  $t = 2$  and  $s_2 = 2$  with  $I_{2,1} = \{1, 2\}$ ,  $I_{2,2} = \{2, 3\}$ . The model with graph  $G^K$  as in Figure 2 (b) has maximum order  $t = 3$ . For  $k = 2$ ,  $s_2 = 3$  with  $I_{2,1} = \{1, 2\}$ ,  $I_{2,2} = \{2, 3\}$  and  $I_{2,3} = \{1, 3\}$ ; for  $k = 3$ ,  $s_3 = 1$  with  $I_{3,1} = \{1, 2, 3\}$ .

Consider  $I \subseteq O$ , let  $\mu_I$  be element of  $\mu_Y$  associated to the entry of the contingency table having value zero for all variables except the variables in  $I$ . Let  $d_I$  be the row of the matrix  $D(\beta)$  corresponding to the first order partial derivative of  $\mu_I$  with respect of  $\beta$ . Note that  $\beta_v$ ,  $v \in K$ , represents the

main effect of the random variable  $A_v$  and for each subset  $I \subseteq O$  such that  $|I| > 1$ ,  $\beta_I$  is the interaction term between the variables in  $I$ . With  $\beta_{\{0,I\}}$  we denote the interaction term between the variables in  $\{0, I\}$ . Moreover,  $\beta_\emptyset = \mu$  is the general mean. With reference to the Example 1, let  $I = \{1, 2\}$ . Then  $\mu_I$  is the expected value of the entry  $(1, 1, 0, 0)$ ,  $d_I$  is the row of  $D(\beta)$  corresponding to the partial derivative of  $\mu_I$  with respect to  $\beta$  and  $\beta_I$  is the term expressing the second order interaction between  $A_1$  and  $A_2$ .

With this notation, to each generic  $ij$ -element of  $D(\beta)$ , we can associate the set  $I$ ,  $I \subseteq O$ , of the observed variables taking value one in row  $i$ ; the set  $J$ ,  $J \subseteq K$ , of variables of which  $\beta_j$  represents either the main effect or the interaction term. Note that both  $I$  and  $J$  could be the empty set. Let  $Z_i$  be the  $i$ -th row of  $Z$ . It is then easy to see that this generic  $ij$ -element is 0 if  $J \not\subseteq I$ ; it is equal to  $e^{Z_i\beta}$  if  $0 \in J$  and  $e^{Z_i\beta} + e^{Z_{i+1}\beta}$ , otherwise. Furthermore, let  $S$  be a complete subgraph of  $G^O$  and  $S' \supset S$ . For  $d_S$  and  $d_{S'}$  and  $\beta_S$  and  $\beta_{\{0,S\}}$  the  $2 \times 2$  square sub-matrix of  $D(\beta)$  has the following structure:

$$\begin{bmatrix} e^a(1 + e^b) & e^{a+b} \\ e^{a+a'}(1 + e^{b+b'}) & e^{a+a'+b+b'} \end{bmatrix} \quad (2)$$

with  $a = \mu + \sum_{I \subseteq S} \beta_I$ ,  $b = \beta_0 + \sum_{I \subseteq S} \beta_{\{0,I\}}$ ,  $a' = \sum_{\{I \subseteq S', I \not\subseteq S\}} \delta(I)\beta_I$  and  $b' = \sum_{\{I \subseteq S', I \not\subseteq S\}} \delta(I)\beta_{\{0,I\}}$ , where  $\delta(I) = 1$  if  $I$  is complete on  $G^O$  and 0 otherwise. Matrix (2) is not full rank if and only if  $b' = 0$ .

We first consider the latent class model, i.e. a model which encodes the following conditional independencies: for each pair of variables  $u, v \in O$ ,  $A_u \perp\!\!\!\perp A_v \mid A_0$ . The following proposition is a restatement of a well-known result.

**PROPOSITION 1.** *A latent class model with  $A_0$  a binary variable is full-rank everywhere in the parameter space if and only if  $n \geq 3$ .*

*Proof.* See [15, 12].

The (unidentified) case of  $n = 2$  is addressed in [5]; see [10, 18, 19, 22] for the Bayesian network literature.

We now remove the assumption that the observed variables are independent conditionally on the latent one, to include a more general class of graphical models  $G^K$  over the  $A_v$  variables,  $v \in K$ , such that  $(0, u) \in E$  for all  $u \in O$ . We first consider graphical models, whose associated complementary graphs  $\bar{G}^O$  are connected and have at least an  $m$ -clique  $C$  where  $m \geq 3$ ; we let  $\bar{C} = O \setminus C$ . The model in Figure 1 is an example of such a

model. The idea is to show that an ordering of the variables in  $\bar{C}$  exists such that, after re-ordering the rows of  $D(\beta)$  accordingly,  $D(\beta)$  admits a lower block-triangular square sub-matrix with full rank. Since  $\bar{G}^O$  is connected for all  $j \in \bar{C}$  there exists a path that connects  $j$  with at least a node of  $C$ .

The following algorithm provides the ordering of the variables in  $\bar{C}$ .

Step 1.  $U \leftarrow \bar{C}$

Step 2.  $T \leftarrow U$  and  $U \leftarrow \emptyset$ .

Step 3. Check if  $T$  is empty, in which case  $\bar{C}$  is ordered; otherwise search the farthest node  $j \in T$  from  $C$ , i.e. with the node with the highest number of edges in the path going from  $j$  to  $C$ ; if there is more than one node choose any one;

Step 4. Let  $J_j$  be the intersection of the set  $T$  and the set of nodes inside the path from  $j$  to  $C$ ; order  $J_j$  starting from  $j$ ; let  $T \leftarrow T \setminus J_j$ .

Step 5. If the last node of  $J_j$  is connected to a node of  $C$ , then append  $J_j$  to  $U$  as the last group of elements (so  $U \leftarrow U \cup J_j$ ); otherwise, if the last node of  $J_j$  is connected to some node  $i \notin T$  order  $J_j$  before  $i$  in  $U$ ; go to Step 3.

LEMMA 1. *Let  $G^K$  be an undirected graphical model over the binary variables  $(A_0, A_1, \dots, A_n)$  with  $A_0$  unobserved and with  $(0, u) \in E$ , for all  $u \in O$ . Assume that in  $\bar{G}^O$  there exists an  $m$ -clique  $C$ ,  $m \geq 3$ . Let  $\bar{C} = \{O \setminus C\}$  and  $M_1$  be the sub-matrix of  $D(\beta)$  formed by the rows  $d_i$  and  $d_{\{i,j\}}$ , with  $i \in \bar{C}$  and  $j$  such that  $(i, j) \in \bar{E}$ , and by the columns  $\beta_i$  and  $\beta_{\{0,i\}}$ . Then  $M_1$  has rank equal to  $2|\bar{C}|$  everywhere in the parameter space if  $\bar{G}^O$  is connected.*

*Proof.* If  $\bar{G}^O$  is connected, there exists an ordering (see the previous algorithm) of the nodes of  $\bar{C}$  such that for any  $i$ ,  $1 \leq i \leq |\bar{C}|$ , the node  $j = i + 1$  is such that  $(i, j) \in \bar{E}$ . Such ordering generates  $|\bar{C}|$  pairs. Let  $M_1^*$  be the sub-matrix of  $M_1$  made up of the rows  $d_i$ ,  $d_{\{i,j\}}$ ,  $1 \leq i \leq |\bar{C}|$ . Then  $M_1^*$  is a  $2|\bar{C}|$ -square lower-block triangular matrix with blocks  $M^i$  associated to row  $d_i$ ,  $d_{\{i,j\}}$ , and columns  $\beta_i$  and  $\beta_{\{0,i\}}$ . The structure of  $M^i$  is as (2) with  $a = \mu + \beta_i$ ,  $b = \beta_0 + \beta_{\{0,i\}}$ ,  $a' = \beta_j$  and  $b' = \beta_{\{0,j\}}$  since by construction  $(i, j) \in \bar{E}$ . As  $\beta_{\{0,j\}} \neq 0$  by the faithfulness assumption, it follows that  $M_1^*$  is full rank and so is  $M_1$ .

◇

LEMMA 2. *Let  $G^K$  be an undirected graphical model over the variables  $(A_0, A_1, \dots, A_n)$  with  $A_0$  unobserved and with  $(0, u) \in E$ , for all  $u \in O$ . Let  $I_{k,r}$  be a complete subgraph of  $G^O$  with  $k \geq 2$ . Suppose that there exists a sequence  $\{I_s\}_{s=0}^{q+1}$ ,  $q \geq 0$ , of complete subgraphs of  $G^O$  such that  $I_0 =$*

$I_{k,r}$ ,  $I_s \neq I_{s'}$ ,  $s \neq s'$ , with  $s, s' \in \{0, \dots, q+1\}$  satisfying the following assumptions:

- (a) for all  $s \in \{0, \dots, q\}$  and for all  $i \in I_s$  there exists  $j \in I_{s+1}$  such that  $(i, j) \notin E$ ;
- (b) for all  $s \in \{0, \dots, q\}$ ,  $|I_s| = k$  and  $|I_{q+1}| < k$ ;

then  $D(\beta)$  contains at least one square sub-matrix  $M_{k,r}$  of order  $2(q+1)$  formed by the rows  $d_{I_s}$  and  $d_{\{V, I_s\}}$ ,  $V \subseteq (I_{s+1} \setminus I_s)$ , and by the columns associated to  $\beta_{I_s}$  and  $\beta_{\{0, I_s\}}$ ,  $s \in \{0, \dots, q\}$ , that has full rank everywhere in the parameter space.

Conversely, if  $D(\beta)$  is full rank everywhere in the parameter space, then for any complete subgraph  $I_{k,r}$  of  $G^O$  with  $k \geq 2$  there is a sequence of complete subgraphs satisfying the assumptions (a) – (b).

*Proof.* We prove the sufficiency first. Consider all the sub-matrices of  $M_{k,r}$ . Observe that a row, and therefore a column, cannot be chosen twice in a  $M_{k,r}$  matrix, as  $I_s \neq I_{s'}$ . By ordering the rows and columns according to the sequence of  $\{I_s\}$ , the matrix  $M_{k,r}$  is seen to be lower block triangular. The blocks are  $N_0, \dots, N_q$  where  $N_s$  is formed by the rows  $d_{I_s}$  and  $d_{\{V, I_s\}}$  and by the columns associated to  $\beta_{I_s}$  and  $\beta_{\{0, I_s\}}$ . Therefore  $N_s$  is as (2). Then,  $\text{rank}(M_{k,r}) = \sum_{s=0}^q \text{rank}(N_s)$  and is full if and only if the blocks are full rank, that is if the rank of each block is equal 2. Suppose that there is an index  $s$  such that no block  $N_s$  has full rank, for all choices of  $V \subseteq I_{s+1}$ . Then, from (2), there exists a strictly positive constant  $0 < h < 1$  such that

$$\begin{cases} e^{\sum_{I \subseteq I_s} (\beta_I + \beta_{\{0, I\}})} = h e^{\sum_{I \subseteq I_s} \beta_I} \left( 1 + e^{\sum_{I \subseteq I_s} \beta_{\{0, I\}}} \right) \\ e^{\sum_{I \subseteq I_s \cup V} \delta(I)(\beta_I + \beta_{\{0, I\}})} = h e^{\sum_{I \subseteq I_s \cup V} \delta(I)\beta_I} \left( 1 + e^{\sum_{I \subseteq I_s \cup V} \delta(I)\beta_{\{0, I\}}} \right) \end{cases} \quad \text{for all } V \subseteq I_{s+1}$$

where  $\delta(I) = 1$  is 1 if  $I$  is complete in  $G^O$  and 0 otherwise. The previous system implies

$$\begin{cases} \sum_{I \subseteq I_s} \beta_{\{0, I\}} = \ln \frac{h}{1-h} \\ \sum_{I \subseteq \{I_s \cup V\}, I \not\subseteq I_s} \delta(I)\beta_{\{0, I\}} = 0 \quad \text{for all } V \subseteq I_{s+1}. \end{cases}$$



From the fact that the model is graphical we obtain:

$$\begin{cases} \sum_{I \subseteq I_s} \beta_{\{0, I\}} = \ln \frac{h}{1-h} \\ \sum_{I \subseteq I_s^V} \beta_{\{0, V, I\}} = 0 \quad \text{for all } V \subseteq I_{s+1} \end{cases} \quad (3)$$

where  $I_s^V = \cap_{j \in V} \{i \in I_s : (i, j) \in E\}$  is the intersection, for varying  $j \in V$ , of the subsets of nodes in  $I_s$  connected to  $j$  in  $G^O$ . Note that by (a) for  $V = I_{s+1}$ ,  $I_s^{I_{s+1}} = \emptyset$ . This implies that  $\beta_{\{0, V\}} = 0$ , which contradicts the faithfulness assumptions since  $I_{s+1}$  is a complete subgraph of  $G^O$ . Therefore there exists for each  $s$  a full rank block  $N_s$  and the square sub-matrix  $M_{k,r}$  is full rank everywhere in the parameter space.

We now prove the necessity. Since  $D(\beta)$  is full rank everywhere, the sub-matrix of  $D(\beta)$  formed by all rows of  $D(\beta)$  and by the columns  $\beta_{I_{k,r}}, \beta_{\{0, I_{k,r}\}}$  is full column rank for all  $\beta \in \Omega$ .

Note that if there exists a sequence of complete subgraphs satisfying (a), but such that for some  $s \in \{1, \dots, q\}$   $|I_s| > k$ , then there exists also a sequence satisfying  $|I_s| = k$ : in fact, if for  $i \in I_s$  there exists a  $j \in I_{s+1}$  such that  $(i, j) \notin E$ , then  $I_{s+1}$  can be chosen in a way that  $|I_{s+1}|$  cannot be greater than  $|I_s|$ . Therefore, either there is no sequence of  $I_s$  such that (a) is satisfied or there is no  $I_{q+1}$  such that  $|I_{q+1}| < k$ .

Going by contradiction, suppose that for  $I_{k,r} = I_0$  there is no complete subgraph  $I_1$  in  $G^O$  such that for each  $i \in I_0$  there is  $j \in I_1$  with  $(i, j) \in E$ . Select the sub-matrix  $C_{k,r}$  formed by the columns  $\beta_{I_0}, \beta_{\{0, I_0\}}$  and all the rows such these two columns have non-zero components, that is select all rows  $d_V$ ,  $V \supseteq I_0$ . (Note that in all other rows the two elements are both 0). Denote with  $\Omega_{k,r} \subset \Omega$  the following subspace:

$$\begin{cases} \sum_{I \subseteq I_s} \beta_{\{0, I\}} = \ln \frac{h}{1-h} \\ \beta_{\{0, V_0\}} + \sum_{I \subseteq I_0^{V_0}} \beta_{\{0, I, V_0\}} = 0 \end{cases} \quad (4)$$

where  $V_0$  is any complete subgraph in  $G^O$  such that for each  $j \in V_0$  there is at least a  $i \in I_0$  with  $(i, j) \in E$  and  $I_0^{V_0} = \cap_{j \in V_0} \{i \in I_0 : (i, j) \in E\}$ . Violation of assumption (a) implies that  $I_0^{V_0} \neq \emptyset$ . Then, it is easy to verify that for  $\beta \in \Omega_{k,r}$  the columns of  $C_{k,r}$  are linearly dependent. In fact, consider system

(4) where  $I_s = I_0$ . System (4) has solution for  $\beta \in \Omega_{k,r}$ . This contradicts the assumption that  $D(\beta)$  is full rank everywhere.

Suppose now that there exists a  $I_0 = I_{k,r}$  such that there is no sequence for  $I_0$  such that  $|I_{q+1}| < k$ . We should find a square sub-matrix  $C_k$  of  $D(\beta)$  with full rank having columns associated to  $\beta_{I_s}$  and  $\beta_{\{0, I_s\}}$ ,  $s \in \{0, \dots, q\}$ , i.e. such that the columns associated to  $\beta_{I_s}$  and  $\beta_{\{0, I_s\}}$  are linearly independent. From the previous derivations, we should consider the rows associated to  $I_s$  and  $V_S = \{I_s, I_{s+1}\}$  (otherwise, the system (4) has solution for  $\beta \in \Omega_{k,s}$ ). But, as there is no  $I_{q+1}$  such that  $|I_{q+1}| < k$ ,  $I_{q+1}$  coincides with some  $I_s$  in the sequence. Therefore, we cannot build a sub-matrix  $C_k$  with full rank.

◇

Obviously, the sequence of complete subgraphs satisfying the assumptions (a) and (b) in Lemma 2 is not necessarily unique.

With reference to Figure 1, let  $I = \{1, 2\}$ . The square sub-matrix with rows  $d_I$  and  $d_{\{4, I\}}$ , and columns  $\beta_I$  and  $\beta_{\{0, I\}}$  is full rank, as the sequence  $I_0 = \{1, 2\}$ ,  $I_1 = \{4\}$  satisfies the assumptions of Lemma 2. Let  $I = \{2, 3\}$ . The square sub-matrix with rows  $d_I$  and  $d_{\{5, I\}}$  and columns  $\beta_I$  and  $\beta_{\{0, I\}}$  is also full rank, as the sequence  $I_0 = \{2, 3\}$ ,  $I_1 = \{5\}$  satisfies the assumptions of Lemma 2. The same holds for  $I_0 = \{3, 4\}$  and  $I_0 = \{4, 5\}$ .

*Remark 1.* If there is a sequence satisfying the assumptions (a) and (b) but  $I_s = I_{s'}$  for some  $s \neq s'$ ,  $s < s'$ , then there is also a shorter sequence satisfying the same assumptions, which is constructed by excluding the interactions from  $I_{s+1}, \dots, I_{s'}$ .

*Remark 2.* An equivalent formulation of the assumption (b) of Lemma 2 is that for  $s \in \{0, \dots, q\}$  and for all  $i \in I_s$  there exists  $j \in I_{s+1}$  such that  $i$  and  $j$  are connected in the complementary graph  $\bar{G}^O$ .

*Remark 3.* The fact that the assumptions (a) – (b) of Lemma 2 hold for all complete subgraphs  $I_{2,r}$  of  $G^O$  with  $|I_{2,r}| = 2$ ,  $r = 1, \dots, s_2$ , does not imply that they hold also for the complete subgraph  $I_{k,v}$ ,  $v = 1, \dots, s_k$  of  $G^O$  such that  $I_{k,v} \supset I_{2,r}$ . As a matter of fact, consider the complementary graph  $\bar{G}^O$  of Figure 3. We can see that conditions (a) – (b) hold for each complete subgraph of  $G^O$  containing two nodes. In particular for  $I_0 = \{1, 4\}$  we have  $I_1 = \{2\}$  and for  $I_0 = \{3, 4\}$ , we have  $I_1 = \{2\}$ . However, the set  $I = \{1, 4, 5\}$  does not verify the condition (b), as  $bd(I) = \{2, 3, 6\}$ , with

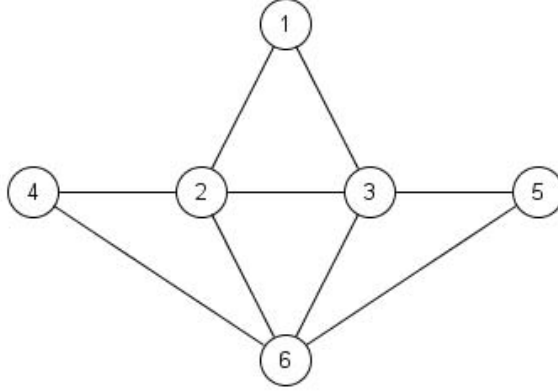


Figure 3: The complementary graph  $\bar{G}^O$  of Remark 3

$\{2, 3, 6\}$  a complete subgraph in  $\bar{G}^O$  with no single node in  $\{2, 3, 6\}$  adjacent to all node in  $I$ .

Suppose that for each fixed order  $k$  of interaction,  $k \in \{2, \dots, t\}$ , the sets  $I_{k,r}$ ,  $r = 1, \dots, s_k$ , satisfy the assumptions of Lemma 2. For each  $I_{k,r}$  then there is a full rank sub-matrix  $M_{k,r}$  of  $D(\beta)$  with rows  $d_{I_s}$   $d_{\{V, I_s\}}$ ,  $V \subseteq I_{s+1}$ , and columns  $\beta_{I_s}$  and  $\beta_{\{0, I_s\}}$ ,  $s \in \{0, \dots, q\}$ . We denote with  $P_k$  the matrix formed by all rows of  $D(\beta)$  and columns used to build all the matrices  $M_{k,r}$ ,  $r \in \{1, \dots, s_k\}$ . By construction, a row, and therefore a column, cannot appear in more than one  $M_{k,r}$ . Then,  $P_k$  is a sub-matrix of  $D(\beta)$  which is full column rank as it is block-triangular matrix with full-rank blocks  $M_{k,r}$ . In fact, the matrix  $P_k$  has zero components in the columns associated to  $\beta_{\{I_{k,r'}\}}$  and  $\beta_{\{0, I_{k,r'}\}}$  for  $r' \neq r$ , so  $P_k$  is a lower block-triangular matrix with blocks full rank everywhere in the parameter space, and is therefore full rank for all  $\beta \in \Omega$ . The following Lemma then holds.

LEMMA 3. Let  $P = [P_2 | \dots | P_t]$  be the sub-matrix of  $D(\beta)$ , with  $P_k$ ,  $k \in \{2, \dots, t\}$ , constructed as previously described. Then  $P$  is full column rank everywhere in the parameter space.

*Proof.* From the fact that the model is graphical,  $P$  is lower block-triangular matrix, as if  $\beta_I = 0$  then  $\beta_{I'} = 0$  for all  $I' \supset I$ . The blocks are full column rank everywhere in the parameter space.

◇

*Remark 4.* The assumptions of Lemma 3 imply that the complementary graph  $\bar{G}^O$  is connected. In fact, let  $\bar{G}^O$  be a complementary graph with two connected components and consider any set formed by two nodes not belonging to the same connected component. A sequence  $\{I_s\}_{s=1}^{q+1}$  with  $|I_{q+1}| = 1$  cannot exist, because the nodes in  $I_s$  (for  $s \in \{0, \dots, q\}$ ) belong to two different components of  $\bar{G}^O$  and the intersection between the neighborhood of  $i, j \in I_q$  in  $\bar{G}^O$  is the empty set.

The next proposition leads to the same conclusion; the proof allows to determine the subspace where the rank of the parametrization is not full.

**PROPOSITION 2.** *Let  $\beta$  be the vector of the parameters of an undirected graphical model  $G^K$  over the binary variables  $(A_0, A_1, \dots, A_n)$ , with  $A_0$  unobserved and  $(0, u) \in E$ , for all  $u \in O$ . If  $\bar{G}^O$  is not connected, then  $D(\beta)$  has no full rank everywhere in the parameter space.*

*Proof.* If  $\bar{G}^O$  is not connected, then  $\bar{G}^O$  has two or more connected components. Let  $\bar{G}^1 = (V_1, E_1)$  and  $\bar{G}^2 = (V_2, E_2)$  be two of them. For any  $i \in V_1$ , consider any complete set  $I_i$  in  $G^O$  of nodes adjacent in  $\bar{G}^O$  to  $i$ . For any  $j \in V_2$ ,  $(i, j) \in E$  and  $(u, j) \in E$  for any  $u \in I_i$ . Let  $\Omega_i \subset \Omega$  be the subspace of  $\Omega$  such that

$$\beta_{\{0,j\}} + \sum_{J \subseteq I_i} \beta_{\{0,J,j\}} = 0. \quad (5)$$

Then, the rank of  $D(\beta)$  is not full for  $\beta \in \Omega_i$ .

◇

We can then prove the following:

**PROPOSITION 3.** *Let  $\beta$  be the vector of the parameters of an undirected graphical model  $G^K$  over the binary variables  $(A_0, A_1, \dots, A_n)$ , with  $A_0$  unobserved and  $(0, u) \in E$ , for all  $u \in O$ . A necessary and sufficient condition for  $D(\beta)$  to be full rank everywhere in the parameter space is that:*

- (i)  $\bar{G}^O$  contains at least one  $m$ -clique  $C$ , with  $m \geq 3$ ;
- (ii) if for each complete subgraph  $I_0$  of  $G^O$ , there exists a sequence  $\{I_s\}_{s=1}^{q+1}$  of complete subgraphs in  $G^O$  such that:

- (a) for  $s \in \{1, \dots, q\}$   $|I_s| = |I_0|$  and  $|I_{q+1}| < |I_0|$ ;
- (b) for  $s \in \{1, \dots, q\}$  and for all  $i \in I_s$  there exists  $j \in I_{s+1}$  such that  $i$  and  $j$  are connected in the complementary graph  $\bar{G}^O$ .

*Proof.* We prove the sufficiency first. Let  $D_C$  be the sub-matrix of  $D(\beta)$  with rows corresponding to the all cells with values zeros for all variables not in  $C$ , and columns  $\mu, \beta_i, \beta_{\{0,i\}}$ ,  $i \in C$ . By (i) and Proposition 1,  $D_C$  is full column rank. Let  $D_{\bar{C}}$  be the sub-matrix of  $D(\beta)$  having rows  $d_i, d_{\{i,j\}}$  and columns  $\beta_i, \beta_{\{0,i\}}$ ,  $i \in \bar{C}$  and  $j$  such that  $(i, j) \in \bar{E}$ . From (ii) and Lemma 2,  $D_{\bar{C}}$  is full column rank. The matrix  $D(\beta)$  can be so written:

$$D(\beta) = \begin{bmatrix} D_C & 0 & 0 \\ B_1 & D_{\bar{C}} & 0 \\ B_2 & B_3 & P \end{bmatrix}$$

where  $B_1$ ,  $B_2$  and  $B_3$  are non-zero matrix (we omit the dimension for brevity), while  $P$  is as in Lemma 3. Therefore,  $D(\beta)$  is full rank everywhere.

To see the necessity, note that  $D(\beta)$  is not full rank only if one of the following matrices  $D_C$ ,  $D_{\bar{C}}$  and  $P$  is not full rank. Proposition 1 implies that assumption (i) is also necessary for  $D_C$  to be full rank. From Lemma 2 and 3, assumption (ii) is also necessary for  $D_{\bar{C}}$  and  $P$  to be full rank for all  $\beta \in \Omega$ .

◇

The following Lemma is a restatement of the assumptions of Proposition 3 in terms of the cliques of the subgraph  $G^O$ .

LEMMA 4. *A restatement of the assumption (ii) of Proposition 3 is the following:*

(ii') *for each clique  $C_0$  in  $G^O$  there exists a sequence in  $G^O$   $\{S_s\}_{s=1}^q$  of complete subgraphs such that*

- (a)  $|S_s| \leq |S_{s-1}|$  for  $s \in \{1, \dots, q-1\}$ ,  $S_0 = C_0$  and  $|S_q| = 1$ ;
- (b) for  $s \in \{1, \dots, q-1\}$  and for all  $i \in S_s$  there exists  $j \in S_{s+1}$  such that  $(i, j) \in \bar{E}$ .

*Proof.* It is immediate to see that (ii) implies (ii)'. The proof of the inverse implication is the following. For  $S = C_0$  it is trivial. For  $S \subset C_0$

consider the following restriction on the sets  $S_1, \dots, S_q$  in the sequence for  $C_0$ : let  $I_0 = S$  and, for  $i \in \{1, \dots, q'\}$ , let  $I_i$  be the set of nodes  $v \in S_i$  such that there exists  $j \in I_{i-1}$  with  $(j, v) \in \bar{E}$  and such that the cardinality of  $I_i$  is not greater than  $|S|$  (see the proof in Lemma 2). The existence of  $I_{q'+1}$  with  $|I_{q'+1}| < |S|$  follows from  $|S_q| = 1$ .

◇

Proposition 3 implies that only the models with connected complementary graph can be identifiable. This contrasts with the condition of globally identifiability in graphical Gaussian models given in [20, 23]. The two conditions coincide only in the case with  $n = 3$  or  $n = 4$ . In this second case, an identified model (under both the discrete and Gaussian distribution) has conditional independence graph as in Figure 2 (a). The model associated to Figure 2 (b) is an underidentified one, as it violates assumption (i) of Proposition 3.

*Example 2.* Consider the model with graphs  $G^K$  and  $\bar{G}^O$  as in Figure 1 (a) and (b). The clique of  $G^O$  are  $\{1, 2\}$ ,  $\{2, 3\}$ ,  $\{3, 4\}$ ,  $\{4, 5\}$ . Setting  $I_0 = \{1, 2\}$  we find that  $I_1 = \{5\}$ , that is  $I_1$  is formed by just one element. The same holds for the other cliques.

*Example 3.* Let the cliques in the graph  $G^O$  are  $C_1 = \{1, 4, 7, 9\}$ ,  $C_2 = \{1, 4, 6, 9\}$ ,  $C_3 = \{1, 4, 6, 8\}$ ,  $C_4 = \{2, 4, 7, 9\}$ ,  $C_5 = \{2, 4, 6, 9\}$ ,  $C_6 = \{2, 4, 6, 8\}$ ,  $C_7 = \{1, 5, 7, 9\}$ ,  $C_8 = \{2, 5, 7, 9\}$ ,  $C_9 = \{3, 5, 8\}$ ,  $C_{10} = \{3, 6, 8\}$ ,  $C_{11} = \{1, 5, 8\}$ ,  $C_{12} = \{2, 5, 8\}$ ,  $C_{13} = \{3, 5, 7\}$ . In Figure 4 the corresponding graph  $\bar{G}^O$  is represented. We can verify from the graph  $\bar{G}^O$  that the assumptions of the Proposition 3 hold. For example, for the interaction term involving  $I_0 = \{1, 5, 8\}$  we have the identifying sequence:  $I_1 = \{2, 4, 9\}$ ,  $I_2 = \{3\}$ . By considering  $I_0 = \{1, 4, 6, 8\}$  we have the identifying sequence  $I_1 = \{3, 7\}$ ,  $I_2 = \{4, 6\}$  and  $I_3 = \{5\}$ .

*Example 4.* In Figure 5 the complementary graph  $\bar{G}^O$  associated to an unidentified model is presented. The condition (ii) of Proposition 3 does not hold for  $\{4, 8\}$  (as well as  $\{4, 9\}$ ,  $\{5, 9\}$ ,  $\{5, 8\}$ ) and any of its superset such as  $\{4, 5, 8, 9\}$ , since the set of nodes connected in  $\bar{G}^O$  to  $\{4\}$  and  $\{8\}$  are  $\{6\}$  and  $\{7\}$ , respectively, and  $(6, 7) \in \bar{E}$ . As a consequence, the interaction among  $I_0 = \{4, 5, 8, 9\}$  of Figure 5 is not identifiable for  $\beta$  in (4), which is so constructed. First find all  $V_0$  sets, which in this case are: for  $V_0 = \{6\}$ , then  $I_0^{\{6\}} = \{8, 9\}$  and for  $V_0 = \{7\}$ , then  $I_0^{\{7\}} = \{4, 5\}$ . The unidentified subspace is therefore the following:

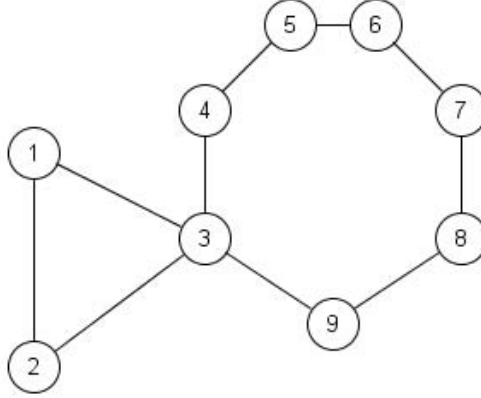


Figure 4: The complementary graph  $\bar{G}^O$  of Example 3

$$\begin{cases} \beta_{\{0,6\}} + \beta_{\{0,6,8\}} + \beta_{\{0,6,9\}} + \beta_{\{0,6,8,9\}} = 0 \\ \beta_{\{0,7\}} + \beta_{\{0,4,7\}} + \beta_{\{0,5,7\}} + \beta_{\{0,4,5,7\}} = 0 \end{cases}$$

The above examples show how to determine the expression of the (sub)space where identifiability breaks down. The following situations may arise:

1. if condition (i) of Proposition 3 does not hold there is no  $\beta^0 \in \Omega$  such that the model is locally identified;
2. if condition (i) of Proposition 3 does hold and condition (ii) of Proposition 3 fails, i.e. there is (at least) a complete subset in  $G^O$  admitting no identifying sequence, the model is locally identified everywhere except in the subspace  $\Omega_{I_0}$  of zero measure for any complete subset  $I_0 \in G^0$  admitting no identifying sequence, with  $\Omega_{I_0}$  given by the following expression:

$$\left\{ \beta_{\{0,r\}} + \sum_{I \subseteq I_0} \delta(r, I) \beta_{\{0,I,r\}} = 0 \quad \text{for any } r \in bd_{\bar{G}^0}(I_0). \right.$$

where  $\delta(r, I) = 1$  if  $\{r, I\}$  is complete in  $G^O$  and 0 otherwise, and  $bd_{\bar{G}^0}(I_0)$  denotes the boundary of  $I_0$  in  $\bar{G}^O$ .

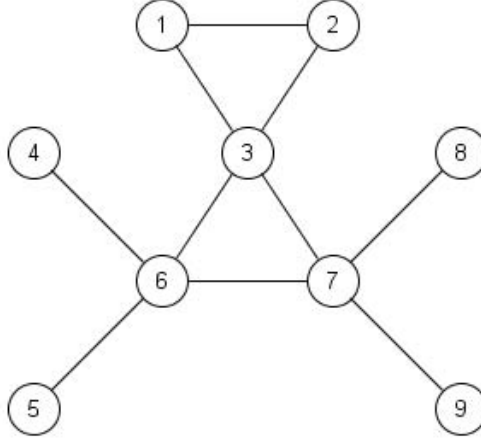


Figure 5: The complementary graph  $\bar{G}^O$  of Example 4

*Example 5.* In Figure 6 the complementary graph  $\bar{G}^O$  associated to an unidentified model is presented. The condition (ii) of Proposition 3 does not hold for  $\{1, 3, 7\}$  and  $\{3, 4, 6, 7\}$ : in fact, in  $\bar{G}^O$  node 3 is connected only to node 5 and node 7 is connected only to node 2, with  $(2, 5) \in \bar{E}$ . By adding the edge  $(3, 7)$  in the missing graph  $\bar{G}^O$ , we get a local identified model.

## 4 Extensions

In this section we extend the condition for local identification to more general models to include (a) observed variables with a generic number of levels and (b) observed variables that are not connected to the unobserved one. The proofs are in the Appendix. The first extension leads to the following:

**THEOREM 1.** *Consider an undirected graphical model  $G^K$  over discrete variables  $(A_0, A_1, \dots, A_n)$ , with  $A_0$  unobserved binary variable and  $(0, u) \in E$ , for all  $u \in O$ . A necessary and sufficient condition for the rank of the parametrization of  $G^K$  to be full everywhere is that conditions (i) and (ii) of Proposition 3 hold.*

As might be anticipated, the graphical model corresponding to Figure



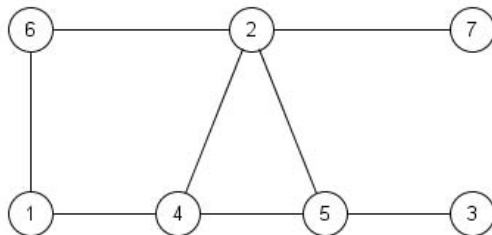


Figure 6: The complementary graph  $\bar{G}^O$  of Example 5

1 is locally identified. Let  $T_1$  be a non-empty set of observed variables not connected to the unobserved one. The idea is to regard the derivative matrix of Proposition 3 as the derivative matrix of this larger model for the rows of  $\mu_Y$  associated to cells with value zero of all variables in  $T_1$ .

**THEOREM 2.** *Consider an undirected graphical model  $G^K$  over discrete variables  $(A_0, A_1, \dots, A_n)$ , with  $A_0$  unobserved binary variable and  $(0, u) \in E$ , for all  $u \in K \setminus \{0 \cup T_1\}$ . Let  $S = \{K \setminus T_1\}$ . A necessary and sufficient condition for the rank of the parametrization to be full everywhere is that the subgraph  $G^S$  satisfies the condition of Proposition 3.*

## 5 Concluding remarks

One of the issues in estimating graphical models with hidden nodes concerns identifiability. In this paper conditions for the identification of discrete undirected graphical models with one hidden binary variable have been determined. For conditionally full rank model, the expression of the unidentified space has been determined, permitting a reparametrization to achieve local identifiability.

The derivations here presented could also be obtained using algebraic techniques, which prove to be an effective tool to study singularities in the parameter space. However, they do not directly provide the interpretation in terms of the associated conditional independence graph, which we believe is the added value of the paper.

Issues of identification of all models that are obtainable as a one to one reparametrization of the discrete undirected graphical model can be ad-

dressed using the results presented in the paper. Extensions to models with more than one latent variable can be addressed exploiting the factorization dictated by the graph. This can be done immediately, provided that the hidden nodes in the graph are isolated. Extensions to models with hidden variables having more than two levels are not straightforward, as the simple structure of the matrix  $D(\beta)$ , as outlined in (2), does not hold any more.

## 6 Appendix

**Proof of Theorem 1:** First assume that all the variables are binary except the  $A_1$  variable which has three levels. Partition  $\beta$  into three subsets  $\beta^a = \{\mu, \beta_0\}$ ,  $\beta^b$  corresponding to the non-zero interaction terms of any order for value in  $\{0, 1\}$  of the observed variables and  $\beta^c$  containing all other parameters. After ordering in a way such that the  $A_1$  variable is running the slowest, the  $D(\beta)$  matrix has the following structure:

$$D(\beta) = \begin{bmatrix} D(\beta^a) & D(\beta^b) & 0_{2^n \times |\beta^c|} \\ D^*(\beta^a) & 0_{2^{(n-1)} \times |\beta^b|} & D^*(\beta^c) \end{bmatrix}$$

where  $[D(\beta^a) \mid D(\beta^b)]$  is the sub-matrix of the derivatives of  $\beta^a$  and  $\beta^b$  and has full rank if conditions (i) and (ii) of Proposition 3 hold. Note that by construction,  $D^*(\beta^c)$  has a similar structure of the sub-matrix of  $D(\beta^b)$  formed by the last  $2^{(n-1)}$  rows and all columns. Therefore  $D^*(\beta^c)$  is full rank if conditions (i) and (ii) of Proposition 3 hold. To see the necessity note that  $D(\beta^b)$  is full rank only if Proposition 3 is verified. Proof of the theorem for  $A_1$  having  $l_v$  levels follows straightforwardly. By a similar argument, extension to a generic number of levels of the  $A_i$  variables,  $i \in O$ , follows.

◇

**Proof of Theorem 2:** Note that  $T_1$  is the set of observed variables such that  $(i, O) \notin E$ . We first focus on models with only binary variables. Let  $T_2 \subseteq S \setminus \{0\}$  be the set of observed variables such that  $(i, j) \in E$ ,  $i \in T_1$ ,  $j \in T_2$ . If  $T_1$  or  $T_2$  is empty the proof is trivial. To start with we assume  $|T_1| = 1$ . Partition  $\beta$  into the subsets  $\beta^d$  containing all the non-zero interaction terms among the variables in  $S$  and  $\beta^e$  containing all the other elements. The non-zero interaction terms among the latent variable and the observed variables are in  $\beta^e$ . The matrix  $D(\beta)'$  has the following structure:

$$D(\beta) = \begin{bmatrix} D(\beta^d) & 0_{2^{|S|-1} \times |\beta^e|} \\ F & D(\beta^e) \end{bmatrix}$$

$D(\beta^d)'$  and  $D(\beta^e)'$  are the derivative sub-matrices for the corresponding elements;  $F$  is a sub-matrix with the same number of rows of  $D(\beta^d)$ . The sub-matrix  $D(\beta^e)$  is full rank because it corresponds to the rank of the design matrix of the model for  $T_1 \cup T_2$ . The sufficiency follows easily from the block-diagonality of the matrix. The necessity follows from the fact that  $F$  has full rank if and only if  $D(\beta^d)$  has full rank. Extension to a generic number of variables in  $T_1$  follows after noting that the matrix  $D(\beta)$  is so built:

$$D(\beta) = \begin{bmatrix} D(\beta^d) & 0_{2^{|S|-1} \times |\beta^e|} \\ F^* & D(\beta^e) \end{bmatrix}$$

where  $D(\beta^e)$  is the derivative sub-matrix for the vector  $\beta^e$  defined as in the previous step.  $D(\beta^d)$  is the derivative sub-matrix for the vector  $\beta^d = \beta \setminus \beta^e$ ;  $F^*$  is a sub-matrix with the same number of rows as  $D(\beta^e)$ . The same considerations as in the previous case hold. Extension to a generic number of levels of the  $A_i$  variables,  $i \in O$ , follows by induction.

◇

## References

- [1] BOWDEN, R. (1973). The Theory of Parametric Identification. *Econometrica* **41** 1069–1074.
- [2] CATCHPOLE, E.A. & MORGAN B.J.T. (1997). Detecting Parameter Redundancy. *Biometrika* **84**, 1, 187–196.
- [3] CATCHPOLE, E.A., MORGAN B.J.T. & FREEMAN, S.N. (1998). Estimation on parameter-redundant models. *Biometrika* **85**, 2, 462–468.
- [4] DARROCH, J.N., SPEED, T.P. (1983). Additive and multiplicative models and interaction. *Annals of Statistics*, 11, 724–738.
- [5] DASGUPTA, A., SELF, S.G. & DAS GUPTA, S. (2007). Non-identifiable parametric probability models and reparametrization. *Journal of Statistical Planning and Inference* **137**, 3380–3393.

- [6] DRTON, M. (2009). Likelihood ratio tests and singularities. *Annals of Statistics* **27**, 2, 979-1012.
- [7] FORMANN, A.K. & KOHLMANN, T. (1998). Structural Latent Class Models. *Sociological Methods Research* **26**, 530-565.
- [8] GARCIA, L.D., STILMAN, M. AND STURMFELS, B. (2005). Algebraic Geometry of Bayesian Newtorks. *Journal of Symbolic Computation* **39**, 331-355.
- [9] GARRETT E.S. & ZEGER, S.L. (2000). Latent class model diagnosis. *Biometrics* **55**, 1055-1067.
- [10] GEIGER, D., HECKERMAN, D., KING, H., & MEEK C. (2001). Stratified Exponential Families: Graphical Models and Model selection. *Annals of Statistics* **29**, 505-529.
- [11] GEIGER, D., MEEK, C. & STURMFELS, B. (2006). On the toric algebra of graphical models. *Annals of Statistics* **34**, no. 3, 1463-1492
- [12] GOODMAN, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 2, 215-231.
- [13] LAURITZEN, S.L. (1996). *Graphical Models*, Oxford University Press: Oxford.
- [14] LINDSAY, B., CLOGG C. & GRECO, J. (1991). Semiparametric Estimation in the Rasch Model and Related Exponential Response Models, including a Simple Latent Class Models for Item Analysis. *Journal of the American Statistical Association* **86**, 96-107.
- [15] MCHUGH, R.B. (1956). Efficient Estimation and Local Identification in Latent Class Analysis. *Psychometrika* **21**, 331-347.
- [16] PACHTER, L. & STURMFELS, B. (2005). *Algebraic Statistics for Computational Biology*. Cambridge University Press, Cambridge.
- [17] ROTHENBERG, T.J. (1971). Identification in Parametric Models. *Econometrica* **39**, 3, 577-591.
- [18] SETTIMI, R. & SMITH, J.Q. (2000). Geometry, moments and conditional independence trees with hidden variables. *Annals of Statistics* **28**, 1179-1205.

- [19] SMITH, J.Q. & CROFT, J. (2003). Bayesian networks for discrete multivariate data: an algebraic approach to inference. *Journal of Multivariate Analysis* **84**, 387–402.
- [20] STANGHELLINI, E. (1997). Identification of a single-factor model using graphical Gaussian rules. *Biometrika* **84** 1, 241-4.
- [21] STANGHELLINI, E.& VAN DER HEIJDEN, P. (2004). A Multiple-Record System Estimation Method that takes Observed and Unobserved Heterogeneity into Account. *Biometrics* **92** 2, 337-350.
- [22] RUSAKOV, D., GEIGER, D. (2005). Asymptotic model selection for naive Bayesian networks. *J. Mach. Learn. Res.*, **6**, 1-35.
- [23] VICARD, P. (2000). On the identification of a single-factor model with correlated residuals. *Biometrika*, **87** , 1, 199-205